PROCEDURAL FLEXIBILITY IN EARLY UNIVERSITY MATHEMATICS

Anne-Maria Ernvall-Hytönen[1], Peter Hästö[2], Heidi Krzywacki[1] & Sara Parikka[1],*
University of Helsinki[1], University of Turku[2]

ABSTRACT

Procedural flexibility refers to the ability to choose efficient strategies for the solution of mathematical tasks. We study procedural flexibility and accuracy when solving linear and quadratic equations and systems of equations at the beginning of university mathematics. We also investigate the viability of the chosen domains for measuring flexibility in the setting of a tri-phase test. We conclude that university students show procedural flexibility and that accuracy and flexibility correlate positively with students' course success. Our results also indicate that we can measure flexibility for both quadratic equations and systems of equations.

INTRODUCTION

This paper concerns procedural flexibility in mathematics, which is the ability to choose the most appropriate strategy for solving a mathematical task efficiently by taking into account task and context characteristics (Heinze et al., 2009; Rittle-Johnson & Star, 2007; Rittle-Johnson & Star, 2008). Mathematical flexibility is an essential part of deeper understanding of strategies and their use in mathematics (Carpenter et al., 1998). The role of mathematical flexibility as an element of mathematical proficiency has gained increasing attention recently both in research (Torbeyns et al., 2009) and in school education (e.g. Common Core State Standards Initiative, 2012). As Heinze et al. (2009) state, students need to learn accurate and adaptive procedures that are reflected also in everyday, working and student life.

Flexibility is treated not as a trait but as a skill that can be improved by solving non-routine problems and by comparing multiple solutions to fairly routine tasks (Cigdem & Yeliz, 2015; Rittle-Johnson et al., 2012; Star & Rittle-Johnson, 2007; Maciejewski & Star, 2016). To promote mathematical flexibility, researchers have investigated the instructional conditions under which it can be enhanced, including exposure to multiple methods (Rittle-Johnson et al., 2012), comparison of worked examples (Newton et al., 2010; Rittle-Johnson & Star, 2007), prompting and direct instruction (Rittle-Johnson & Star, 2008) and the opportunity to collaborate with peers (Mercier & Higgins, 2013). Researchers have also investigated the extent to which low-achieving students or students with learning difficulties

* All authors contributed equally to this article and are listed in alphabetical order.

can develop and exhibit flexibility (Newton et al., 2010; Newton et al., 2014; Peters et al., 2014; Verschaffel et al., 2007).

This study concerns procedural flexibility at the beginning of university mathematics studies when solving algebraic tasks. Prior research on flexibility has been conducted mostly within arithmetic and elementary algebra among lower and upper secondary students (Blöte et al., 2001; Schneider et al., 2011; Xu et al., 2017). At the university level, mathematical flexibility has been studied in calculus (Maciejewski & Star, 2016) as well as within arithmetics (Shaw et al., 2020). For example, Shaw et al. (2020) found college students to show less flexibility than expected and it seemed, in addition, to be unrelated to cognitive and affective aspects, such as mathematical anxiety, mathematical identity, need for cognition and working memory. Little is still known about flexibility in different mathematical areas and the way it could be investigated at university level despite its essential role in mathematical expertise.

PROCEDURAL FLEXIBILITY

There are different ways to conceptualize mathematical flexibility. Mathematical flexibility has been considered as one of five strands of mathematical knowledge alongside conceptual knowledge and procedural fluency (Kilpatrick et al., 2001). Baroody (2003) considers flexibility as "adaptive expertise" that is seen as the integration of conceptual and procedural knowledge, whereas Star (2007) suggests strategic flexibility is deep procedural knowledge. Some scholars have shown a correlation between measures of conceptual and procedural knowledge and flexibility, although the direction of causation is not clear (McMullen et al., 2017; Rittle-Johnson et al., 2012; Schneider et al., 2011). We approach flexibility as deep procedural knowledge following Star (2007) and thus, we mean by *flexibility* such strategic flexibility that is also called procedural flexibility by some scholars. Specifically, flexibility is considered knowledge of various strategies and ability to choose the most appropriate one to solve a particular task.

The relationship between flexibility and accuracy has emerged in research, for instance, when focusing on development of flexibility among struggling students (Newton & Lynch, 2010). Accuracy is seen as a facet of procedural knowledge, namely the correct application of a viable solution strategy to a given task (Star, 2007). Knowledge of multiple strategies is related to greater accuracy and greater conceptual knowledge in arithmetic (Carpenter et al., 1998) and algebra (Rittle-Johnson & Star, 2007). In this study, we consider accuracy as the ability to solve the task correctly, i.e. to present correct intermediate steps and to arrive at the correct final answer.

Flexibility can be studied with a tri-phase test based on tasks, which can be solved by a well-known standard strategy as well as by a computationally more efficient "situational strategy" (see Methods for more information on the test). Such a test has been used to study flexibility of middle- and high-school students from China, Finland, Spain and Sweden (Xu et al., 2017; Liu et al., 2018; Hästö et al., 2019; Star et al., 2022). Earlier studies found relatively low levels of flexibility, quite high levels of accuracy and a correlation between the two (Xu et al., 2017; Lui et al., 2018; Star et al., 2022; Shaw et al., 2020).

The study has been guided by two research foci of which the first focus is on the way accuracy and flexibility appear in elementary algebra at university level. This focus concerns how viable quadratic equations and systems of equations are as domains for measuring flexibility in tri-phase tests:

1) How accurate and flexible are university students in solving elementary algebra tasks?

The second focus explores the relationship between flexibility and accuracy, and any correlation of these with the course exam:

2a) Is there a connection between flexibility and accuracy among university students in solving algebra tasks?

2b) Is there a connection between flexibility and success in a course exam?

METHODS

Participants

We invited students in a large introductory mathematics course in a Finnish university to participate voluntarily in the study. The course served both mathematics majors and students from other programs, such as computer science, statistics and economics. One of the authors was the lecturer of the course. Data was collected in the beginning of the course in September 2021 via a test that was implemented in the Moodle environment. The online test made it possible for all students to participate despite the special study arrangements due to the pandemic. After completing the test, students were asked to give permission to use their anonymised test results in the study. Students were asked separately to give consent to the use of the scores of the course exam. Completing the test or giving consent to participate in the study had no impact on the course evaluation.

Originally, data was gathered from 125 students but the data from 4 participants could not be used in the analysis due to technical difficulties in entering or extracting data from Moodle.

Data gathering

The test consisted of two parts, a basic part and a tri-phase part for examining flexibility. The basic part included four tasks on arithmetic and algebraic simplification and equation solving (see Table 1) that measured accuracy and also served as a warm-up. Students were asked to give only the final answer for the first three tasks and select the right option in Task 4.

The tri-phase part followed the structure laid out by Xu et al. (2017). In the three phases, participants are asked (1) to solve a task, (2) to provide additional solutions based on different solution strategies and (3) to mark which of their solutions they consider the best. The original test comprising 12 linear equations was completed with paper and pencil in class. Our tri-phase part included only 4 tasks that were delivered with the instructions via Moodle. In contrast to previous studies using tri-phase tests, we did not have any time limits for completing the test.

Table 1. Basic part of the test.

#	Task	Answer
1	Simplify $\frac{2}{5}-\frac{1}{2}$	$-\frac{1}{10}$
2	Simplify $\frac{2/3}{1/5}$	$\frac{10}{3}$
3	Simplify $(2x)^5(-4x)^{-1}$	$-8x^4$
4	Which of the following statements applies to the equation $2(x-1)=3x-(x+1)$? a. The equation has exactly one solution b. The equation has no solutions c. Every real number satisfies the equation d. None of the the above	b.

In phase 1, participants were instructed to solve all four tasks and write the solutions in the leftmost column of a 4×3 grid on a sheet of paper. In phase 2, new instructions appeared and students were asked to provide up to two additional solutions to each task in the remaining columns. Finally, in phase 3, participants were asked to circle for each task the solution they felt was best. Students returned their answers by taking a photograph of the paper and uploading it to Moodle.

The tri-phase part started with Tasks 5 and 6 that were linear equations from Star et al. (2022) and it continued with Tasks 7 and 8, designed for this study, that concerned quadratic equations and systems of equations. The tasks are presented in the first column of Table 2.

We studied flexibility using a tri-phase procedure that has been found to have good psychometric properties and measures both on flexibility and accuracy (Xu et al., 2017). There was some novelties in the way we applied the test. First, we included quadratic equations and systems of equations alongside linear equations. Systems of equations were chosen because of their importance in university mathematics while also being part of the pre-university curriculum, both at lower and upper secondary level. Second, the tri-phase test was administered online, via Moodle environment, instead of face-to-face and consisted of fewer tasks than previous tri-phase studies.

In addition to the test data, we used the score of the final course exam and tested whether course achievement relates to flexibility and accuracy. The course exam consisted of three tasks which test general conceptual understanding related to the main ideas of the course, namely induction, functions and relations as well as polynomial equation solving in complex numbers.

The coding process

The coding was carried out jointly by two of the authors. For tasks 1-4, we started coding the students solutions as either correct (score 1) or incorrect (score 0) due to there being no intermediate steps. The accuracy score of these tasks we labeled as basAcc.

In the tri-phase part of the test, every task was coded separately for accuracy and for strategy. In addition, the solution marked as "best" was recorded. Both the final answer and all intermediate steps of phase 1 had to be mathematically valid in Tasks 5-8 to be considered correct; each task was coded either incorrect (score 0) or correct (score 1). The *accuracy* score for the tri-phase part was determined by the number of correct solutions in phase 1. Correctness in phase 2 was not considered, as per the coding scheme of the tri-phase test established by Xu et al. (2017). The accuracy score of tasks 5-8 we labeled as Acc.

The solution in the tri-phase part was coded as standard strategy (S), situationally appropriate (situational, M), or other (O). The standard and situational strategies were defined as follows (see Table 2 for an overview) and all the other solutions were coded as other strategies. Table 2 shows examples of such authentic solutions.

In task 5, if the first step of the solution of the equation $4(x + \frac{3}{5}) = 12$ was to distribute the 4 on the left side, then strategy was considered standard. If the first step was to divide the equation by 4, then strategy was considered situational. Figure 1 shows an example of solutions of which the first solution (in phase 1) is coded as a standard strategy (S) and the second solution (in phase 2) as a situationally appropriate (M), which has been chosen as "best". In addition, the phase 1 solution is correct and thus the score for accuracy is 1.

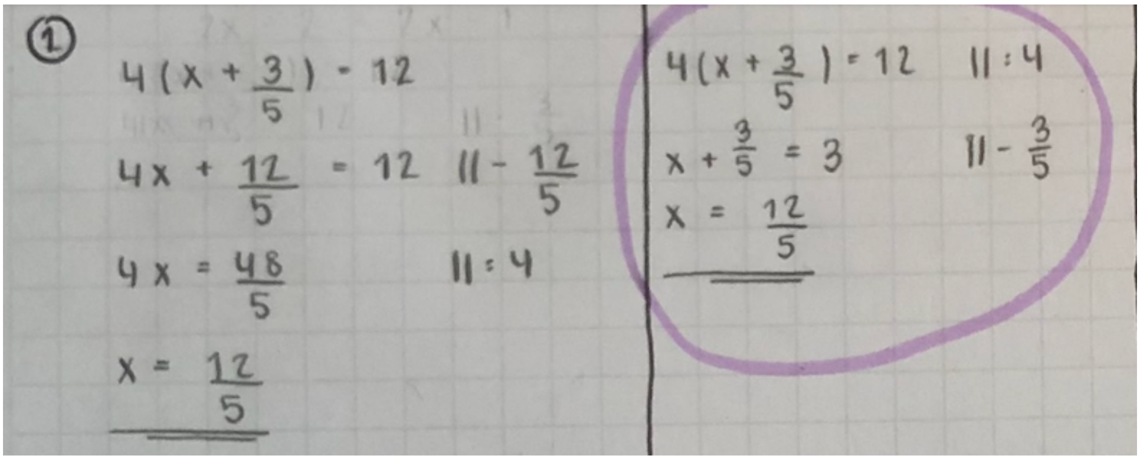

Figure 1. An example of an answer to Task 5.

In task 6, a standard strategy for $5\left(x + \frac{3}{7}\right) + 3\left(x + \frac{3}{7}\right) = 16$ was to start with distributing the 5 and the 3. If the first step was to combine the terms on the left side, resulting in $8\left(x + \frac{3}{7}\right) = 16$, it was coded as a situational strategy.

The standard strategy to solve Task 7, $2x^2 + 4x = 0$, was to start with the quadratic formula. If the first step was to factorize as $x(2x + 4) = 0$ and to use zero-product property, then it was coded as a situational strategy and marked as M. Some students divided the equation with x or $2x$ and thus, came to a conclusion

Table 2. Tri-phase part of the test.

Task	Standard strategy	Situational strategy	Other strategy
5. $\mathbf{4\left(x+\frac{3}{5}\right)=12}$	$4x+\frac{12}{5}=12$ $4x=12-\frac{12}{5}$ $4x=\frac{48}{5}$ $x=12/5$	$x+\frac{3}{5}=3$ $x=3-\frac{3}{5}$ $x=\frac{12}{5}$	$\frac{4\left(x+\frac{3}{5}\right)}{12}=1$ $\frac{x+\frac{3}{5}}{3}=1$ $x+\frac{3}{5}=3$ $x=\frac{12}{5}$
6. $\mathbf{5\left(x+\frac{3}{7}\right)+3\left(x+\frac{3}{7}\right)=16}$	$5x+\frac{15}{7}+3x+\frac{9}{7}=16$ $8x+\frac{24}{7}=16$ $8x=16-\frac{24}{7}$ $8x=\frac{88}{7}$ $x=\frac{11}{7}$	$8\left(x+\frac{24}{7}\right)=16$ $8x=16-\frac{24}{7}$ $8x=\frac{88}{7}$ $x=\frac{11}{7}$	$5+3=\frac{16}{x+3/7}$ $\frac{8}{16}=\frac{1}{x+3/7}$ $\frac{1}{2}=\frac{1}{x+3/7}$ $x=2-\frac{3}{7}$ $x=\frac{11}{7}$
7. $\mathbf{2x^2+4x=0}$	$x=\frac{-4\pm\sqrt{4^2-4\cdot 2\cdot 0}}{2\cdot 2}$ $x=\frac{-4\pm 4}{4}$ $x_1=0, x_2=-2$	$x(2x+4)=0$ $x_1=0, 2x+4=0$ $2x=-4, x_2=-2$	$x^2+2x+1=1$ $(x+1)^2=1$ $(x+1)^2-1=0$ $x_1=0, x_2=-2$
8. $\begin{cases}\mathbf{3x+4y=2}\\ \mathbf{y=2x+7}\end{cases}$	$\begin{cases}3x+4y=2 & \lvert\cdot 2\\ y=2x+7 & \lvert\cdot 3\end{cases}$ $\begin{cases}6x+8y=4\\ -6x+3y=21\end{cases}$ $11y=25, y=\frac{25}{11}$ $\frac{25}{11}=2x+7$ $x=-\frac{26}{11}$	$3x+4(2x+7)=2$ $3x+8x+28=2$ $11x=-26, x=-\frac{26}{11}$ $y=2\cdot\left(-\frac{26}{11}\right)+7$ $y=\frac{25}{11}$	Graphical solution

that either $x=0$ or $x+2=0$, and it was coded as a variant of the situational strategy (code M2). Figure 2 shows an example of a solution that is coded as M2. The solution is incorrect and thus, the score for accuracy is zero as the solution $x=0$ is missing.

Task 8 required the most involved coding since both the subtraction method and the substitution method could lead to being coded as standard or situational. The situational strategy (M) was to substitute $2x+7$ in place of y in the first equation. A variant of the situational strategy (marked as M2) was to multiply the second equation by 4 and subtract it from the first one. The solution was coded as a

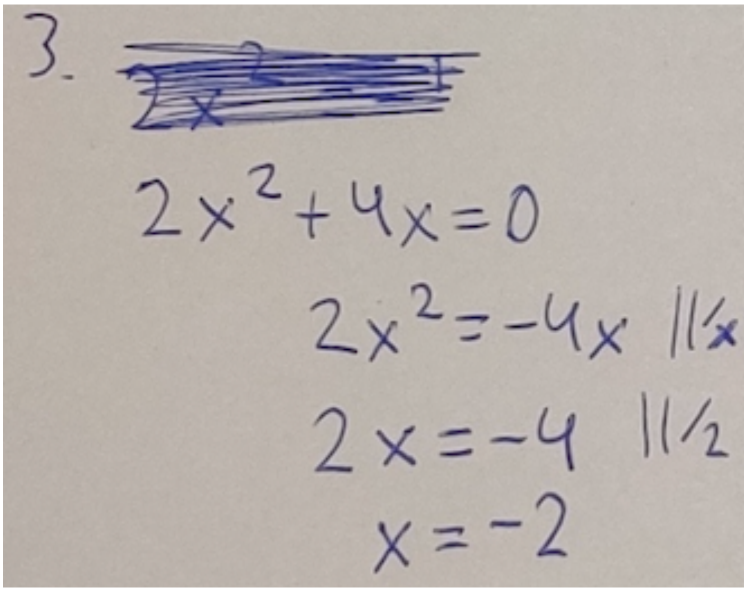

Figure 2. An example of an answer to Task 7.

standard strategy (S) if the equations were multiplied with non-optimal real numbers and then added together. The variant S2 was used for solutions with matrices and S3 for solutions where x was solved from one equation and then substituted or y was solved in a nonoptimal way, for example, solving y from the first equation and then substituting in the second one.

We started our data analysis using the same measures of flexibility and accuracy as Star et al. (2022). Thus, we first merged different variants of the standard strategy (S, S2 and S3) into one category and did the same with the situational strategies (M and M2). A task was deemed flexibly solved if three criteria were fulfilled:

1. The task was solved using a standard strategy.
2. The task was solved using a situational strategy.
3. The strategy chosen as best was a situational one.

The number of flexibly solved tasks gave each participant a flexibility score between 0 and 4, which we labeled as Flex.

Over the course of our analysis, we observed that these criteria led to unreasonable results if only two situational strategies (M and M2) were presented. Therefore, we redefined flexibility to include also the case with two situational strategies without giving a standard strategy. Note that this did not impact flexibility in the linear equations (Tasks 5 and 6), since there is only one type of situational strategy in this context.

The purpose of these changes in the original tri-phase test was to try to measure flexibility in a broader context beyond linear equations. This means that the flexibility scores are not directly comparable to those of earlier studies using tri-phase tests such as Star et al. (2022). A better comparison could be achieved by only using tasks 5 and 6 and comparing them to the corresponding tasks in earlier studies. However, the context of giving instructions in Moodle and not having a time-limit might impact the absolute scores. It should be noted that the time limit was not much of a factor in the original study by Xu et al. (2017) which validated the test, but became an issue only in later studies with non-Chinese participants (Star et al., 2022). In any case, comparison with previous studies is not the main objective of this study.

Statistical data analysis

We used descriptive statistics including frequencies for a basic understanding of our observations. We used the Cronbach alpha coefficient to investigate whether the flexibility in Tasks 5-8 form a consistent scale. The maximum value for alpha is 1, which indicates complete concurrence of the measures. Nunnally (1978) recommends minimum values of 0.7 for early stage research and 0.8 for applied research. To investigate whether the four tasks 5-8 measured the same construct, we also calculated the correlations between the flexibility scores in these tasks. Since the scores are dichotomous (i.e., 0 or 1), tetrachoric correlation is an appropriate measure (Ekström, 2011). The interpretation of the tetrachoric correlation is the same as for the more familiar Pearson correlation: a score of 1 means alignment, 0 means no relationship and -1 means an inverse relationship.

The independent-samples t-test was used to compare the mean scores of different groups on various variables including accuracy and flexibility. We used Pearson correlations to measure the relationship between accuracy, flexibility and exam score. The relation is further studied with a linear regression model predicting the exam score based on flexibility and accuracy.

RESULTS

This study focuses on flexibility when solving algebraic tasks at the beginning of university mathematics. We describe our results by discussing separately flexibility as well as accuracy and thereafter, showing the findings about the relation of flexibility to accuracy and the success in the course exam.

Of the 121 students who participated in the test, 19 did not complete the tri-phase part. Comparing the score on the basic part (tasks 1–4) indicates that these 19 students had marginally lower average scores ($t(119) = 1.962, p = 0.052$). Thus, the students who chose to participate in the entire test seem to be somewhat self-selected. The 19 students are not included in the following analyses so the study is based on data from the 102 students who completed the entire test.

As can be seen in Table 3, all tasks were completed with rather high accuracy. The lowest accuracy scores were in Task 3 (56 %), Task 7 (82 %) and Task 8 (67 %) while the accuracy in the remaining tasks was 90–95 %. The most important reasons for the low accuracy scores were minor calculation errors. In task 7 that was on a second-degree equation, many students forgot to show the second solution. In particular, in approach M2 they divided the equation by x and forgot to take into account the case $x = 0$. In task 8 that was on a system of equations, a typical mistake was to solve only one variable, for instance, solving x correctly but not solving y.

The average number of solutions was just under 2 for each of the tri-phase tasks. For tasks 5 and 6, the standard strategy was well-known (97–98 %) and the situational strategy was quite well-known, whereas the situation for tasks 7 and 8 was the opposite. Only 47 % presented the standard strategy while 94 % used the situational in Task 8. We found that many participants presented two variants of the situationally appropriate strategies (coding M and M2) in Task 8 and, to a lesser extent, in Task 7. This is natural, since there is not as clear a standard solution for Task 7 as for Task 8.

Table 3a. Overview of test results in basic part.

Basic part	Abbreviation	Task 1	Task 2	Task 3	Task 4
Accuracy	basAcc	91 %	92 %	56 %	90 %

Table 3b. Overview of test results in tri-phase part.

Tri-phase part	Abbreviation	Task 5	Task 6	Task 7	Task 8
Accuracy	Acc	95 %	93 %	82 %	65 %
#Solutions		1.98	1.84	1.91	1.83
Standard		98 %	97 %	64 %	47 %
Situational		80 %	73 %	81 %	94 %
Flexibility	Flex	71 %	62 %	43 %	54 %

We calculated the tetrachoric correlations between the flexibility scores in tasks 5–8. As can be seen in Table 4, the pairwise correlations are very high, although the correlations with Task 7 are somewhat lower. This indicates that these questions measure the same concept. The Chronbach alpha of the scale consisting of the four tasks' flexibility equals 0.782. This falls between Nunnally's (1978) recommended thresholds of 0.7 for early stage research and 0.8 for applied research. This also supports these questions measuring the same concept.

Table 4. Tetrachoric correlation of flexibility in Tasks 5–8 (n=102)

	T5	T6	T7	T8
T5	1.000	0.883	0.761	0.807
T6	0.883	1.000	0.633	0.753
T7	0.761	0.633	1.000	0.504
T8	0.807	0.753	0.504	1.000

We also compared the accuracy and flexibility of the students with their exam results. Of the 102 students who completed the whole test, 57 allowed the use of their exam scores for the research whereas 45 did not. The former subgroup had higher Acc, ($t(69.68) = -2.240$, $p = 0.036$, variances not assumed equal) and Flex, $t(100) = 2.039$,$p = 0.044$ and no significant differences in basAcc, $t(100) =$

1.640, $p = 0.104$. Thus it seems that again the 57 students who chose to permit the use of the grade are not wholly representative of all students taking the test.

Figure 3 indicates that most students for each flexibility score are at the high end of the exam score (15–18 points). The 8 students with zero points did not participate in the exam and are excluded from further analysis. Furthermore, the three isolated points (in red) had an undue influence on results such as group variance and so they were excluded from further analyses as outliers.

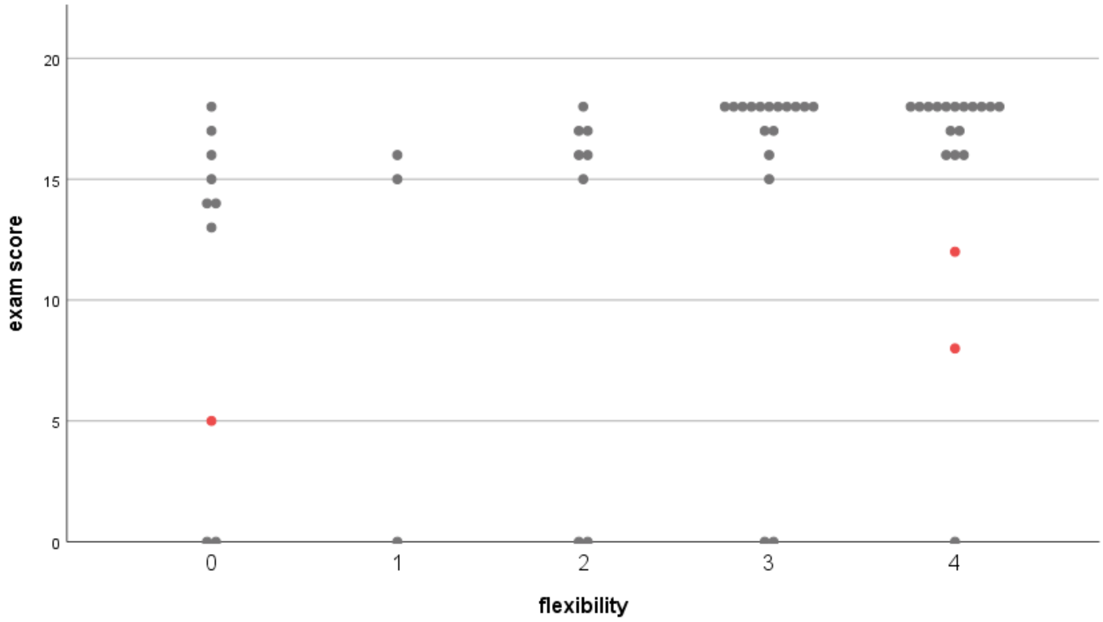

Figure 3. Exam score by flexibility (n=57).

There is a slight trend towards higher exam scores with increasing flexibility scores as indicated by the correlations shown in Table 5. The correlations between basAcc, Acc and Flex are mostly not significant in this group ($n = 46$) whereas the correlations in the whole dataset $n = 102$ are significant at the 0.01-level.

Table 5. Pearson correlations between variables for all participants (left, $n = 102$) and the subset whose exam scores could be used (right, $n = 46$). Asterisks indicate significance at the 0.05-level (*) and the 0.01-level (**).

All	basAcc	Acc	Flex
basAcc	1	0.281**	0.285**
Acc	0.281**	1	0.379**
Flex	0.285**	0.379**	1

Subset	basAcc	Acc	Flex	exam
basAcc	1	0.312*	0.150	0.386**
Acc	0.312*	1	0.282	0.416**
Flex	0.150	0.282	1	0.616**
exam	0.386**	0.416**	0.616**	1

A linear regression model predicting exam score based on basAcc, Acc and Flex was significant $F(3,42) = 13.9, p < 0.001$, with an adjusted R^2-value of 0.463. BasAcc and Flex were statistically significant predictors, accounting for 28 % and 6 % of the variance, respectively. The scatter plot in Figure 3 suggests that there was a ceiling effect with the most common score being the maximum of 18 when the flexibility score was 3 or 4. A stronger correlation might result if the exam had more power to distinguish high-achieving students.

DISCUSSION

We have studied students' flexibility in elementary algebra at the beginning of their university studies. Recall that by flexibility we mean strategic flexibility following Star (2007). In regard to the first research question, we found high levels of accuracy (65–95 %) and quite high levels of flexibility (43–71 %) over the four tasks. As expected, these are higher than for Finnish 11th grade high-school students in the advanced mathematics track, who had 66 % accuracy and 36 % flexibility on average in all 12 tasks in the longer version of the tri-phase test with time-limit (Star et al., 2022). Note that only two tasks of the original test were used in our study (as tasks 5 and 6) and neither study is based on representative samples. Still, the similar but somewhat higher numbers can be considered a validating factor for the new, shorter test presented in this article.

Our tri-phase test included two new tasks designed for this study in addition to the two tasks from previous studies (e.g., Liu et al., 2019; Star et al., 2022). We derived measures of accuracy and flexibility following the model of previous research using the tri-phase test (Xu et al., 2017). In terms of flexibility, this means that the participant must produce both a standard and a situationally appropriate solution for the situation, and mark the latter as "best". However, we found that for task 8 (system of equations), this requirement was not as clear due to the lack of an obvious standard algorithm. Students are typically taught to solve systems of equations by three different means (addition method, substitution method and graphical method) without prioritizing any of them as a standard method. Furthermore, each method can be applied in various ways, e.g. by choosing which variable to solve first. Many participants produced two distinct solutions both of which we considered situationally appropriate. To account for this, we considered a new, broader flexibility variable which also included the case of two situational solutions with either chosen as "best". The same applied also to task 7 (quadratic equation), albeit to a much lesser extent. This demonstrates the difficulty of applying the operationalization of flexibility from outside the domain of linear equations, which have a very clear standard algorithm (Xu et al., 2017; Hästö et al., 2019; Liu et al., 2019; Star et al., 2022).

Based on our study, we conclude that quadratic equations and systems of equations seem to be viable domains for measuring flexibility with a tri-phase setup but more work is needed for producing a better, more reliable measure. The new flexibility measure performed well in many respects in this study. The correlations between the flexibility in different tasks were high and the reliability of the flexibility scale as measured with Chronbach's alpha was good for early stage research. There is still a need in future investigations to improve on this, for instance by refining the task instructions or by adding more tasks.

In the tri-phase test, one measures both flexibility and accuracy with the same tasks. These two measures correlate quite strongly in most groups that have been studied (Star et al., 2022). With the new setup, we compared flexibility with accuracy not only based on the tri-phase test data but also the score from an exam taken several weeks after the tri-phase test. The correlation between accuracy and flexibility at 0.379 was lower than for Finnish students in middle and high-school (Star et al., 2022). This may be due to a ceiling effect in the accuracy variable of the tri-phase test. Interestingly, the correlation with the exam score was higher, especially for flexibility. Indeed, we found that flexibility in the tri-phase test was a much better predictor of exam score than accuracy, accounting for 29 % of the variance, while accuracy was not statistically significant. This contrasts with the results of Hästö et al. (2019) that accuracy in the tri-phase test was a better predictor than flexibility of exam results in the national matriculation examination taken one year after the test. This may reflect the higher correlation between accuracy and flexibility in high-school students and the ceiling effect, or it may mean that the course exam measures mathematical skills where flexibility is of greater value when compared to the national exam at the end of high-school.

In response to the second research question, we found that accuracy and flexibility were correlated, but not as strongly as in previous studies. Moreover, we can say quite confidently that flexibility's connection with accuracy in the tri-phase test differed from its connection with the exam score. However, contrary to expectation the latter was stronger rather than weaker. This, again, may be due to a ceiling effect in the test's accuracy measure.

LIMITATIONS AND FUTURE STUDIES

Our results are based on a convenience sample from an introductory course in one university. This limits the generalizability to other populations. We had 121 participants, but many did not return a full set of data. After removing participants with incomplete data, those who did not consent to the use of their exam score, and a few outliers, only 46 participants remained for some analyses. Furthermore, each stage of self-selection may result in a group which is slightly more flexible and mathematically skilled. Thus the results may give too rosy a picture of the flexibility of university students. This self-selection may also be responsible in part for the ceiling effect that we observed both in the accuracy measurement in the test as well as in the exam. It is unclear how this affects the correlation between flexibility and accuracy, as it might either increase or decrease it.

Administering the tri-phase test through the Moodle platform proved largely successful and it seems possible to continue developing the tri-phase test in this context. Improving the formulation of the questions as well as motivating the students to answer all questions could increase the completion rate. Furthermore, the new tasks on the quadratic equation and the system of equations can be further developed to fit better with the tri-phase framework and thus increase the reliability of the test. The resulting Moodle-based tri-phase test can then be scaled up to a larger study with a more representative sample.

REFERENCES

Baroody, A.J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody, & A. Dowker (Eds.), *The Development of arithmetic concepts and skills: constructive adaptive expertise* (pp. 1–33). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410607218

Blöte, A., Burg, E., & Klein, A. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, *93*, 627–638. https://doi.org/10.1037/0022-0663.93.3.627

Carpenter, T.P., Franke, M.L., Jacobs, V.R., Fennema, E., & Empson, S.B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education, 29*(1), 3–20. https://doi.org/10.2307/749715

Cigdem, A., & Yeliz, Y. (2015). Common and flexible use of mathematical non routine problem solving strategies. *American Journal of Educational Research, 3*, 1519–1523. https://doi.org/10.12691/education-3-12-6

Common Core State Standards Initiative (2012). *The Standards: Mathematics*. http://www.corestandards.org/the-standards/mathematics. Accessed August 25, 2022.

Ekström, J. (2011). The phi-coefficient, the tetrachoric correlation coefficient, and the Pearson-yule debate. *Department of Statistics, UCLA*. https://escholarship.org/uc/item/7qp4604r. Accessed August 29, 2022.

Heinze, A., Star, J.R., & Verschaffel, L. (2009) Flexible and adaptive use of strategies and representations in mathematics education. *ZDM - Mathematics Education, 41*, 535–540. https://doi.org/10.1007/s11858-009-0214-4

Hästö, P., Palkki, R., Tuomela, D., & Star, J. R. (2019). Relationship between mathematical flexibility and success in national examinations. *European Journal of Science and Mathematics Education, 7*(1), 1–13. https://doi.org/10.30935/scimath/9530

Kilpatrick, J., Swafford, J. O., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Liu, R.D., Wang, J., Star, J.R., Zhen, R., Jiang, R.H., & Fu, X.C. (2018). Turning Potential flexibility into flexible performance: moderating effect of self-efficacy and use of flexible cognition. *Frontiers in Psychology, 9*, 646 https://doi.org/10.3389/fpsyg.2018.00646

Maciejewski, W., & Star, J.R. (2016). Developing flexible procedural knowledge in undergraduate calculus. *Research in Mathematics Education, 18*, 1–18. https://doi.org/10.1080/14794802.2016.1148626

Mercier, E., & Higgins, S. (2013). Collaborative learning with multi-touch technology: Developing adaptive expertise. *Learning and Instruction, 25*, 13–23. https://doi.org/10.1016/j.learninstruc.2012.10.004

McMullen, J., Brezovszky, B., Hannula-Sormunen, M., Veermans, K., Rodríguez-Aflecht, G., Pongsakdi, N., & Lehtinen, E. (2017). Adaptive number knowledge and its relation to arithmetic and pre-algebra knowledge. *Learning and Instruction, 49,* 178-187. https://doi.org/10.1016/j.learninstruc.2017.02.001

Newton, K., & Lynch, K. (2010). Understanding the development of flexibility in struggling algebra students. *Mathematical Thinking and Learning, 12*. https://doi.org/10.1080/10986065.2010.482150

Newton, K., Willard, C., & Teufel, C. (2014). An examination of the ways that students with learning disabilities solve fraction computation problems. *The Elementary School Journal, 115*, 1-21. https://doi.org/10.1086/676949

Nunnally, J.C. (1978). *Psychometric theory (2nd ed.)*. New York: McGraw-Hill.

Peters, G., De Smedt, B., Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2014). Subtraction by addition in children with mathematical learning disabilities. *Learning and Instruction, 30*, 1-8. https://doi.org/10.1016/j.learninstruc.2013.11.001

Rittle-Johnson, B., & Star, J.R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., Star, J.R. (2008) Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, *18*(6), 565–579. https://doi.org/10.1016/j.learninstruc.2007.09.018

Rittle-Johnson, B., Star J.R., & Durkin, K. (2012). Developing procedural flexibility: are novices prepared to learn from comparing procedures? *Journal of Educational Psychology, 82*(3), 436–55. https://doi.org/10.1111/j.2044-8279.2011.02037.x

Schneider, M., Rittle-Johnson, B., & Star J.R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology Journal*, *47*(6), 1525–1538. https://doi.org/10.1037/a0024997

Shaw, S.T., Pogossian, A.A., & Ramirez, G. (2020). The mathematical flexibility of college students: The role of cognitive and affective factors. *British Journal of Educational Psychology, 90*(4), 981–996. https://doi.org/10.1111/bjep.12340

Star, J.R., Tuomela, D., Joglar-Prieto, N., Hästö, P., Palkki, R., Abánades, M.Á., Pejlare, J., Jiang, R.H., Li, L., & Liu, R.D. (2022). Exploring students' procedural flexibility in three countries. *International Journal of STEM Education, 9*, 4. https://doi.org/10.1186/s40594-021-00322-y

Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Solving subtractions adaptively by means of indirect addition: Influence of task, subject, and instructional factors. *Mediterranean Journal for Research in Mathematics Education*, *8*(2), 1–30. https://doi.org/10.1080/10986060802583998

Verschaffel, L., Torbeyns, J., De Smedt, B., Luwel, K. & Van Dooren, W. (2007). Strategy flexibility in children with low achievement in mathematics. *Educational and Child Psychology*, *24*, 16-27.

Xu, L., Liu, R.D., Star, J.R, Wang, J., Liu, Y. & Zhen, R. (2017). Measures of potential flexibility and practical flexibility in equation solving. *Frontiers in Psychology, 8*, 1368. https://doi.org/10.3389/fpsyg.2017.01368